Tehran University of Medical Sciences
Global Strategies and International Affairs (GSIA)

Office of Vice President for Research Affairs

Research Project/Thesis Proposal Form

Research Title:

Application of Machine learning to predict conversion of MCI to AD

Full Name of the Project Manager(s):

Ali Fele Paranj, Narjes Hajimollaheydar, Zahra Amjadi, Tauseeq Fatimah,Mahdie Karimzade , Abolfazl Torabi , Sina Khezri , Mahdieh Esmaeili , Mahshad Fadaei

School/Research Center:

Students'Scientific Reseach Center (SSRC)

Type of Research:

Basic ☐ Applied/Clinical ■ H S R ☐

This research is a:

Student Thesis ■

TUMS Research Project ☐

Joint Research Project ☐

Project Description

Tehran University of
Medical Sciences
International Campus

If a thesis, specify the level: Undergraduate ■ M.Sc. ☐ MPH☐ PhD☐ Postdoc. ☐

Type of Study: Please mark

	Case series
	Cross sectional
	Case / control
	Cohort
	interventional clinical trial /
*	Experimental
	Pharmaceutical Study
	Implementation of a scientific/ executive Method
	Test Review
	Method Review
	Qualitative
'	Health System Management Study
	Software Design

Information about the Project Manager(s)

- Full Name(s):
- Academic Rank:
- School/Research Center:
- Department:
- Research Location:
- Expected Duration:
- Current Position and work location:
- Work Phone Number:
- Work Address:
- E-mail Address:
- **Contact Number in case of emergency:**

Research Project Team: (Other supervisors, advisors, students, other partners)

1	Full Name	Position and Academic Rank	Type of Involvement	E-mail address and Phone Number	Partner's Signature
2	Zahra amjadi	M.D. Psychiatrist			
3	Fatimah Tauseeq	Medical student			
4	Sina Khezri	Medical student			
5	Mahdie karimzade	Medical student			
6	Mahdieh Esmaeili	Computer Software Engineer			
7	Mahshad Fadaei	B.A. of psychology			
8	Ali Fele Paranj	BSc of physics			
9	Abolfazl Torabi	Medical student			
10	Narjes Hajimollaheydar	BSc of physics			
11					

1- Proposal Abstract (max 300 words):

Alzheimer's disease is a progressive neurodegenerative illness known by the effect on memory and cognitive functions, which has a substantial emotional and financial burden on patients and their families and health care systems. AD is the first leading cause of dementia globally with no definite treatment , An estimated 5.7 million Americans have Alzheimer's dementia. Total payments in 2018 for health care and hospice services for people with the age of 65 years with dementia are estimated to be $277 billion. Conventional treatments for AD only reduce the progression, which highlights the importance of early diagnosis. Mild cognitive impairment (MCI) is a concept by which a person has mild symptoms of cognition abilities but still can carry out daily tasks independently. This stage prodromes AD. Early diagnosis of MCI and proper treatment is an essential step toward decreasing AD's load on society. A mathematical model estimates that early and accurate diagnosis could save up to $7.9 trillion in medical and care costs. Machine Learning(ML) is a very promising era of scientific study to analyze large amounts of data and gaining the ability to judge upon the probable outcome of new subject. ML is a branch of artificial intelligence applies for regression and classification purposes. The trained machine with classifier algorithms will use data features to make calculation of the most probable outcome. This project aims to use ML in order to analyze data(features) such as MRI studies, biomarkers in blood and CSF, and demographic information and cognitive performance tests obtained from Alzheimer's Disease Neuroimaging Initiative cohort (ADNI) aiming to predict the transformation of MCI to probable AD in the near future. Another valuable advantage of this technique is highlighting the most valuable data(features) in order to predict the chance of AD from MCI in a specific and unfamiliar subject.

Key Words :

Alzheimer’s Disease, ADNI , Machine Learning , Dementia , Mild cognitive impairment

2- Rationale and Backgrounds:

Alzheimer's disease (AD) is a common progressive neurodegenerative disorder that is the most prevalent type of dementia in elderly people. As the life expectancy increases, the number of AD patients will also increase correspondingly, resulting in a substantial socio-economic burden [1, 2] Despite the high burden of this disease, preventive and the therapeutic interventions for AD have largely failed. [3, 4]. As the progression of the neuropathology in AD starts years in advance before clinical symptoms of the disease become apparent and progressive neurodegeneration has irreversibly damaged the brain, emerging treatments will likely have the greatest effect when provided at the earliest disease stages. Thus, the prompt identification of subjects at high risk for conversion to AD is of great importance [5]. The progression of AD comprises a long, unnoticed preclinical stage, followed by a prodromal stage of Mild Cognitive Impairment (MCI) [6]MCI divided to two groups included progressive MCI (MCI-p)and stable MCI(MCI-s). MCI-p are Individuals who progressed to AD during the follow-up and Individuals who did not have a change of diagnosis and remained stable during the follow up time are belong to MCI-p group[7] Since early detection and classification of the disease can help treat this disease in its early stages, many studies have been performed to predict Alzheimer's disease. These studies mainly suggest the advantage of using multivariate analysis over univariate techniques as they account for the relationship between variables and are less prone to classification errors[8, 9] A growing number of studies have been using machine learning (ML) and multivariate analysis methods to classify individuals at risk of progression to AD[5, 7, 10] which provide tools to analyze information and observe inherent disease-related patterns in the data .One of the advantages of these classifiers is the potential use for detecting AD at the prodromal stages, before clinical manifestation.[11] Wee et al. in 2012 proposed a new approach of extracting morphological information from structural MRI images and they constructed a regional cortical thickness similarity map for each subject to describe the correlative changes in cortical thicknesses between pairs of different regions of interest (ROIs)and obtained an accuracy of 92.4% (sensitivity and specificity of 90.4% and 94.3%) for classification of AD and healthy control subjects(CTL) and an accuracy of 75.1% (sensitivity and specificity of 63.5% and 84.4%) for classification of converters and non-converters MCI subjects within 36 months. These results were achieved by using a multi-kernel SVM method and an integration of ROI-based and correlative morphological MRI features from 509 AD, MCI, and CTL subjects.[12]and Chincarini et al proposed a fully automated technique to extract discriminative features based on selected pathology-specific volumes of interest in order to compute a classification index. They assessed the accuracy of the classification index on prediction of conversion from MCI to AD within a time frame of 2 years. The authors used a Random Forest algorithm to reduce the dimensionality of the feature set and select the most relevant and important features for classification. Subsequently a support vector machines (SVM) classifier was used to compute the classification index. The authors investigated the performance of the proposed algorithm on a population of 635 AD, MCI, and CTL subjects from the ADNI cohort. An accuracy of 92% was reported for classification of AD and CTL subjects, with an accuracy of 68% for MCI prediction [13].these two studies are based on MRI features but some studies using multivariate analysis and do study based on multi-modality features. Recently Wolz et al have proposed a framework based-on manifold learning to extract features from imaging modalities and combine them with non-imaging metadata. The result is a unified biomarker that can be used for data analysis and visualization. The authors investigated the performance of the proposed method on classification of 420 AD, MCI, and CTL subjects from the ADNI cohort. MRI was employed as the imaging modality and CSF biomarkers and APOE genotype were used as non-imaging metadata. A linear SVM classifier was utilized to distinguish different diagnostic groups. The classification of AD versus CTL subjects using all imaging and non-imaging data resulted in an accuracy of 88% (sensitivity and specificity of 85% and 91%), and the classification of MCI converters versus stable MCI subjects resulted in an accuracy of 69% (sensitivity and specificity of 68% and 70%).[14]and Kohannim et al. investigated the discriminative power of different biomarkers for classification of AD subjects and prediction of MCI conversion at one-year follow-up. A linear SVM classifier was employed to discriminate different groups of subjects using the following data: MRI, FDG-PET, CSF biomarkers, APOE genotype, age, gender, and body mass index. Numerical summary measures of hippocampal, ventricular, and temporal lobe volumes were used as MRI features. In addition, a numerical summary based on a predefined temporal lobe ROI was used as a FDG-PET feature. The total dataset included 635 AD, MCI, and CTL subjects from the ADNI cohort; however, three different subsets were defined according to data availability of each modality. The best AD versus CTL classification accuracy was 90.7%,which was obtained using

MRI,FDG-PET, and CSF biomarkers.[15] A series of recent studies use machine learning and deep learning approach .Angela tam et al in 2019 proposed algorithm to predict 3-year conversion to AD in MCI subjects, based on a weighted rank average ensemble of several supervised machine learning algorithms. 550 samples from ADNI open database were used with a 3 year follow up. The final algorithm developed, consisted of predictors such as sociodemographic and clinical characteristics along with neuropsychological test scores and also several supervised machine learning algorithm. Results indicated an AUROC of 0.88 , sensitivity of 77.7%, specificity of 79.9% on excluded test data. The specificity of algorithm was 40.2% for 100% sensitivity[5] Garam Lee et al in 2019 conducted his research for predicting conversion from MCI to probable AD by applying a deep learning approach, multimodal recurrent neural network. For achieving this, cross-sectional neuroimaging biomarkers in the baseline and also longitudinal cerebrospinal fluid (CSF) and cognitive performance biomarkers are obtained from the Alzheimer's Disease Neuroimaging Initiative cohort (ADNI). In this study a total of 1,618 ADNI participants aged 55 to 91 were used, which include 415 cognitively normal older adult controls (CN), 865 MCI (307 MCI converter and 558 MCI non-converter), and 338 AD patients. The results of this research showed when using only single modality of data separately, prediction model for MCI conversion to AD yielded up to 75% accuracy (area under the curve (AUC)=0.83) and by incorporating longitudinal multi-domain data, the prediction model achieved the best performance with 81% accuracy (AUC=0.86).[16] Chin-Fu Liu et al in 2019 applied a deep learning framework which utilize Siamese neural networks trained on paired lateral inter-hemispheric regions in order to harness the discriminative power of whole-brain volumetric asymmetry. Subjects and scans in this study were selected from the ADNI 1, ADNI GO, ADNI 2, and BIOCARD databases .In total, 3566 1.5T T1 scans across 819 subjects were selected from the ADNI database; and 744 1.5T T1 scans across 324 subjects were selected from the BIOCARD database all scans were processed and parcellated by MRICliud. 274 volumetric features were searched by explicitly learning the asymmetry encoding in Siamese NN which in comparison with other deep learning methods provides comparable prediction balanced accuracies (BACC) 0.9436 for ADNI and 0.9220 for combined ADNI and BIOCARD.[17] Ali Ezzati et al in 2019 conducted his research by applying demographics and structural MRI measures for classification of cognitively normal (CN) vs Alzheimer's disease (AD) participants from ADNI and subsequently applied the model with the best classification performance to participants with MCI to predict AD conversion in different follow up time up to 4 years.424 CN ,249 AD individuals and 656 MCI patients were used from ADNI database. The performance of classifiers (decision trees , support vector machine , K-nearest neighbor , ensemble linear discriminant, boosted trees and random forest) ranged from 80% to 92 %. [7]

3- Research Objectives and Goals:

A: Main Objective:

To make differentiation between subjects with mild cognitive impairment, Alzheimer Dementia Type (DAT) and normal cognitive; also to establish a prediction method of the rate of the progression of the MCI group of patients to AD with lower error-1 type (false positive).

B: Specific Objectives:

To Get access to the database of the Alzheimer disease neuroimaging initiative cohort by presenting the proposal within one week of submission based on the ADNI website statement.
To create a set of preprocessed data from MRI PET EEG blood and CSF biomarkers and demographic information consisting of sex, age and academic level of the subjects produced either by the intrinsic researchers or the operators of ADNI cohort applicable by the machine learning.
To gather information arising from the measures mentioned in methodology including hippocampal size, cortical thickness, APOE4 and etc.
To train the “machine” via train data within a period of 6 months with respect to processing power by one of the colleague expert in the discipline of physics using CNN, RNN and decision tree due to the sort of data
To Make a classification of introduced subjects in three grouped marked as normal cognitive, mild cognitive impairment and Alzheimer dementia type using test data.
To analyze output information of classification using statistical methods.
To make the prediction of the rate of progression of MIC to AD conversion by the created method.

C: Goals:

improvement of estimation accuracy of MCI to AD conversion probability.

4- Research Questions and Hypothesis:

One of the most valuable goals of scientific research on Alzheimer's Disease (AD) is to develop methods to diagnose it in it is most primal stages in order to stop or decelerate its progression. This goal could be achieved by investigating AD from lots of different aspects. For example, how the disease progresses in a specific age group or what tests or data could predict the outcome more accurately or even how many subjects should be investigated to produce a legitimate conclusion and so on.
Tackling this issue from all these aspects needs a large amount of time and resources if done by human individuals. On the other hand, integrating all these data to produce a logical answer is very complex task or may even seem impossible.
What may be a solution to solve this issue?
Computers and in particular Artificial Intelligence (AI) provide a very sensible solution since lots of these tasks could be done in matters of hours if not minutes since they are much more efficient in computing huge amount of calculations than humans.
The rational procedure to create an AI which is specialized in achieving this goal is to feed data from known cases of AD or its prodromal stages in order to create a pattern. A pattern by which the disease progresses throughout the course of time and by utilizing the constructed pattern by AI we may be able to predict the most probable outcome for a certain subject In particular stage of disease in near future. The mentioned process has been used through lots of experiments in past each presenting different efficiency and result. The following step would be promoting the accuracy of the proposed pattern. There are several logical strategies to exalt the accuracy of the proposed algorithms.
One is to increase the number of subjects investigated by the algorithm. Another would be to expand the number of data analyzed on each subject, which in this case are data such as imaging modalities (MRI, PET scans), Cognitive function tests (MMSE, memory, language, etc.), Demographic (Age, Sex, educational state), Genetics. Another strategy is following these subjects in temporal context in order to make a rational conclusion.
The core idea behind this study is that by combining all the aforementioned strategies plus maximizing the precision of Machine Learning (a certain branch of AI being used in this study) algorithms on performing each strategy, the illustrated pattern grows in accuracy. Hence working in favor of achieving the primary goal the research which is to diagnose AD in earliest stages.
In the first step Gathering a large group of AD cases and Mild Cognitive Impairment (prodromal stage to AD) and subjects with no cognitive problems (Normal Cognitive) in order to make sensible comparison between them. At the same time running various imaging or functional tests on each subject and validating all by the same sets of standards adds to the problem even more. We aim to solve this obstacle by using data from ADNI which is basically a valuable directory of AD patients, MCI s and NC (control group) which have been followed over the course of 3 to 5 years and have lots valuable test data.
The initial step for machine is to classify a given subject in one the following classes Normal Cognitive (NC), stable MCI (s MCI), progressive MCI (p MCI), AD. This step is achieved by reviewing large amount of data (MRI, PET scans, genetics, cognitive function tests, etc.) on each specific class and creating a distinctive line between them via supervised training of the machine. In details the machine is basically thought by a human supervisor how different combination of test data classify each subject in a predetermined class of subjects by ADNI. The next step is to give a presumption on how fast disease is growing in a progressive MCI subject which is prone to AD but still hasn't developed certain symptoms and signs and ultimately may have any capacity for prevention.
This objective is met by selecting a certain group of subjects in ADNI which provide information on disease progressing gradually over time ultimately to construct a more justifiable pattern of disease. This goal is planned to be obtained by selecting a group of subjects which at baseline were diagnosed to be p-MCI and have gradually transformed to AD at some point in the cohort plus have test data recorded in several points along (called target group). The detailed description of such selection is discussed in projects method in details.
Given that ADNI has acquired test data form each subject in equivalent time periods, extracting test data from ADNI in form of numerical figures allows us to follow the pattern by which they change over time and enabling to construct the next step in this process. The constructed data is analyzed by the trained classifier to see if it fits in AD category or not if not this process is repeated multiple times till it fits AD

class, then based on the number of cycles gone forward we can argue how long it takes for each individual to produce AD class features. Then by comparing our constructed features to target group data we can validate the proposed strategy in accuracy.

5- Research Design and Methods:

the machine learning (ML) methods had been used vastly in solving the regression and classification problems in recent years and Python as a powerfull programming language had been a very useful tool for these purposes.

I. Study design:

In our study, we will use the ML methods using python with Alzheimer's disease data set to both classify the subjects and estimate the MCI to AD progress. The subjects can be classified into four different groups: 1. NC, 2. progressive MCI, 3. None-progressive MCI and 4. AD.

II. Subjects :

Inclusion/exclusion criteria

Inclusion Criteria:
All subjects are selected through ADNI

Exclusion criteria:
Chronic traumatic encephalopathy , Age-associated memory impairment, Alcohol or drug abuse, Depression, Vitamin B12 deficiency, Cerebrovascular disease (and vascular dementia) , Hearing or visual impairment, Hypernatremia, Hypernatremia, Hypothyroidism or hyperthyroidism, Normal pressure hydrocephalus, Parkinson's disease with dementia, Polypharmacy, Volume depletion, Wernicke-Korsakoff Syndrome, Aphasia, Carotid Artery Stenosis Imaging, Chronic Myelogenous Leukemia (CML), Cortical Basal Ganglionic Degeneration, Dementia in Motor Neuron Disease, Lewy Body Dementia, Depression, Frontotemporal Dementia and Frontotemporal Lobar Degeneration Huntington, Disease Dementia, Normal Pressure Hydrocephalus, Parkinson Disease, Parkinson-Plus Syndromes, Parkinson-Plus Syndromes, Prion-Related Diseases, Vascular Dementia, Wilson Disease , factious subjects, Hypertension, Cardiovascular disease, Head Traumatic injury, Brain lesions, any other severe medical conditions affecting brain

Sampling

To make this possible, different ML algorithms will be used. The data set which is going to be used is the ADNI which provides different variety of data sets including MRI images, PET scan images, biomarkers, cognitive tests, and the demographic data.

Recruitment plans

All subjects are recruited by ADNI

Method of assignment to study groups

Done By ADNI

III. Data collection :

Variables: outcomes, predictors, confounders

All the variables are stated in Variables Table below

Measures/instruments

Microsoft Excel, Python

Procedures

In our study, we aim to run the ML methods in Python using Alzheimer's disease data set to both classify the subjects and estimate the MCI to AD progress.
The first step is classifying the subjects into four different groups: 1. NC, 2. progressive MCI, 3. None-progressive MCI, and 4. AD.
Our first goal is to classify the subject into one of the mentioned groups. Then if the subject is in the progressive group, we try to estimate the MCI to AD progress.
To make this possible, different ML algorithms will be used. The data set which is going to be used is the ADNI, which provides different variety of data sets, including MRI images, PET scan images, biomarkers, cognitive tests, and the demographic data.
MRI and PET scan images are 2D images. MRI contains the structural, and PET scan includes the functional data of the brain. Biomarker's data set includes the level of tau-protein and beta-amyloid in CSF.
And the demographic data set contains sex, age, education level of the patients.
For the classification problem, we examine deep learning methods including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and other classic machine learning methods like K Nearest Neighbors (KNN), Randomized KNN, Support Vector Machine Classifiers (SVM), Naive Bayes, Decision Tree, and Random Forest will also be used. After training the machine with these algorithms, ensemble learning methods will be used to gain better results (that in our study is to have lower False-positive score).
We will use different methods for different data sets we get from ADNI. As mentioned before, ADNI will provide MRI images, PET scans, biomarkers, cognitive tests, and demographic data.
We have two data sets. One is raw data acquired from ADNI and the other one Processed Data (PD) data set, which produced by processing raw data. Some examples of raw data are the volume of the brain regions, the thickness of the cortex, Hippocampus size, regional function levels of the brain regions, etc. There are many useful softwares developed for these purposes. Some of them are: Analysis of Functional Neuroimages (AFNI), FMRIB Software Library (FSL), Computerized Anatomical Reconstruction Toolkit (CARET), FreeSurfer, Statistical parametric mapping (SPM). We utilize suitable ones for generating the PD data set.

CNN method will be used with PET and MRI images to classify the subjects. The CNN algorithm has some convolutional layers in it, and by using these layers, the algorithm can 'see' the image and get trained by them to classify the subjects.

In parallel with that, the classic machine learning methods will be used on the Processed Data (PD).

Recurrent Neural Networks are the best in classifying data in the form of time series. This method also can be used to predict the value in the next time step. So the RNN will be performed on the time sequence like data such as MRI and biomarkers and cognitive tests to classify the subjects. After the classification, for those who are in the MCI group, a trained RNN will be used for progress estimate purposes.
The CRNN, which is the combination of CNN and RNN, will be tested on MRI images taken at different times to profit both RNN and CNN benefits.
Other pseudo heuristic methods will be tested during this study. For example, the sequence of MRI images with 3-month steps is in the time domain. Converting it to the frequency domain using Fourier Transform can tell us about the frequencies in the time series. Now the frequencies can be used as the training data set of classic machine learning algorithms,
We have two major approaches to the progress estimate.
The first one is using the sequence estimator methods like RNN to estimate the next statues given the previous ones. Each time we estimate the following status, the estimated status can be classified using the classifiers. Now in each step, we can measure the accuracy of classifiers, and as it reaches a threshold, we can regard the status as converged

to AD. The number of sequence estimators now can be regarded as the point at which an MCI turns to AD. To make it clear, suppose that the data set of MRI images in ADNI are taken with 3-month steps. And the status of an imaginary subject will be AD after 4 sequence estimates. So, we can estimate that after 12 months (4 middle estimates * 3-month step for each), the status of the subject turns to AD.
The other approach which needs to be more transparent in the meantime, is using the encoders. The perceptron encoders that are widely used in feature reduction and word to vector purposes can be used to turn our sequential data to a data set with universal time labels.
We also may use other statistical analysis tools to analyze the data set. The methods mentioned so far were all Machine Learning based. So we need to bring other tools to make the study more accurate. For example, one possibility is finding the Kramers-Moyal coefficients of the data set. These coefficients contain critical information. For example, the first coefficient (which is called the drift term) can tell us the best prediction about the future value of the stochastic variable. Like other statistical predictions, our prediction contains an error. The second coefficient (which is called the diffusion) tells us about the error. As another example of the advanced statistical analysis, by statistical tools, it is possible to find the potential of a random process. The local minimums of the potential will give us the attractors of the random process. In our example, the attractors are the different states of the AD that a patient can be in.

The effectiveness of the mentioned methods on our specific study is not apparent and deducted after evaluating the methods on the data set to each other. So this report of the method contains our best guesses that may work.

IV. Intervention:

No intervention is made in data sets

V. Statistical considerations:

Sample size

for classification use all subjects in ADNI will be used

for estimation phase a group of about 200 MCI subjects are deduced to be eligible based on previous studies

Data analysis

We use several combination of Machine Learning algorithms like deep learning algorithms and dense perceptron's on dataset ADNI with 1242 person from 55 to 90 year that the number of normal, MCI and AD's people is sequential 360,644 and 238. The ADNI dataset has around 2042 FDG-PET scans and 2402 MRI images including all the longitudinal time-points by using the diversity of algorithms. The RNN algorithm use for biomarker, cognitive test and MRI and the CNN method is used for MRI and PET scan, the CRNN use for period of image during of MRI. The classic methods like Decision Trees, Random Forests, KNN, Randomized KNN, Support Vector Machine Classifiers, Naive Bayes used for get high accuracy of learning Machine of data test. We measured the mean accuracy rates (percentage of epochs correctly classified) of the models obtained for each feature combination in the test sets.

6- Variables Table:

	Variable	Definition	Qualitative		Quantitative	Variable			Measurement method	scale
			Ordinal	Nominal	Discrete	Continuous	Independent	Dependent	Measurement Method	Scale
1	Demographic info	Age				*	*			second
		Sex		*			*			
		Education level	*				*			
		Martial state		*			*			
2	Cognitive functions	Evaluation of memory, orientation, judgment, problem solving, attention, concentration, naming, repetition, comprehension, language etc.			*		*		MMSE, ADAS-Cog13, CDR etc.	
3	CSF measurement	Aβ1-42				*	*		lumbar puncture	mg/μl
		t-tau				*	*			
		p-tau181				*	*			
4	Genetic and Genomic Analyses	APOE		*			*			
5	MRI	hippocampal volume				*	*			mm^3
		Cortical thickness				*	*			mm
		Fimbria volume				*	*			mm^3

7- Estimated total time to complete the research (in months):

8- Research Timeline Table:

Prepare a list of the activities planned for the research proposed. Mark with X the appropriate cells to reflect the time (each cell represents one month) and duration of each activity. An example of activities is provided in the first three rows.

	Activities	Duration of the activity																																		
	Example Activity 1				X	X	X	X	X	X	X	X																								
	Example Activity 2							X	X	X	X	X	X	X	X																					
	Example Activity 3									X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X											
1	**Data extraction and evaluation**	X	X	X	X																															
2	**Preprocessing**					X	X	X	X																											
3	**Generating numeric data**									X	X																									
4	**Building models**											X	X	X	X	X	X	X	X	X																
5	**Training the models**																				X	X														
6	**Evaluating the models**																						X	X	X	X	X									
7	**Statistical analysis**																											X	X	X	X	X				
8	**General evaluation**																																X	X		
9																																				
10																																				
11																																				
12																																				

9- Ethics: (Attach Ethical Consent Form if applies)

Data compiled for this proposal was collected through the online available dataset i.e., “Alzheimer’s Disease Neuroimaging Initiative 3” (ADNI3). Consent was gained from all participants by ADNI, and the details of the study were made transparent to all the partakers. Also, the data gathered through ADNI was in accordance with Good Clinical Practice (GCP) guidelines, and fully conformed to the regulations for the “Protection of Human Subject Research”, codified in 45 CFR part 46, Protection of Human Subjects , 21 CFR part 50 and part 56, Institutional Review Board (IRB) and /or the International Conference on Harmonization (ICH) E6 , Health Insurance Portability and Accountability Act (HIPAA) , State and Federal regulations apart from the remainder of pertinent laws and requirements.

10- Safety Considerations:

No actual participants were recruited by the authors of this article, instead, all the data provided was collected from the online database i.e., Alzheimer's Disease Neuroimaging Initiative 3 (ADNI3). During the data obtainment, all ADNI's safety protocols were followed. Some examples of safety parameters include symptom checklists, vital signs, physical and neurological examination, laboratory tests, etc. Data Safety Monitoring Board of ADNI Coordination Center was also involved throughout the study the ensure that the safety concordance. The data was reviewed by this body on a quarterly basis. Efforts were made to quickly identify and deal with adverse effects, for example, life-threatening situations, nosocomial outcomes among others.

11 - Limitations:

the limitations of our study can be divided into two major categories: 1. Computational 2. dataset for the first one the main limitations are on CPU clock speed and the amount of the available RAM for the second one the limitation is the size of dataset. As everybody may know, for an accurate machine learning project, a big data set is needed. but the size of the available data on the ADNI is small. Particularly, for the 'prediction' part of our study we will use the timeseries data which is recorded during the cohort. The problem with this study is that the depth of the time series (number of values measured over the cohort) is very small (maximum 10 records for different data sets). Some (like PET scan) even does not have time series type data and other limitation are such as:

1.ADNI is that our population represents a clinical trial population and not an epidemiologically selected real-life population.

2.Our subjects do not include those with cortical strokes, cancer, heart failure, substance abuse, etc. Therefore, the extent to which the results from ADNI can be generalized to the entire population remains to be determined.

3.ADNI only studies subjects aged 55–90 years, and there is considerable evidence that AD pathology may begin to occur in the human brain well before this age.

4.A forth limitation of ADNI is the type of data that are not being collected including computerized neuropsychological testing, electroencephalogram, magnetoencephalography, magnetic resonance spectroscopy, metabolic and inflammatory markers, and lifestyle information.

5.An inherent limitation of this and other pattern classification studies using the ADNI dataset is the reliance on the clinical diagnosis of AD as the "ground truth" (gold standard). The clinical diagnosis of probable AD has an accuracy of 70–90%relative to the pathological diagnosis.

6.Another limitation of this study is the relatively short follow-up period of three years. Although the development of prognostic models for long-term dementia prediction is warranted, short-term dementia prediction can be useful for selecting high-risk MCI patients in clinical trials

7.Our sample size is small and the small sample size makes it difficult to build an effective model

8. Finally, although in this study we considered only the APOE genotype as a generic predictor of progression, genome wide association studies have been used to identify several other genes that likely contribute to the development of AD.

12 - References:

1. Bagattini, C., et al., Predicting Alzheimer's disease severity by means of TMS–EEG coregistration. Neurobiology of aging, 2019. 80: p. 38-45.
2. Fan, Y., et al., Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. Neuroimage, 2008. 41(2): p. 277-285.
3. Rahimi, J. and G.G. Kovacs, Prevalence of mixed pathologies in the aging brain. Alzheimer's research & therapy, 2014. 6(9): p. 82.
4. Schneider, J.A., et al., Mixed brain pathologies account for most dementia cases in community-dwelling older persons. Neurology, 2007. 69(24): p. 2197-2204.
5. Grassi, M., et al., A Novel Ensemble-Based Machine Learning Algorithm to Predict the Conversion From Mild Cognitive Impairment to Alzheimer's Disease Using Socio-Demographic Characteristics, Clinical Information, and Neuropsychological Measures. Frontiers in neurology, 2019. 10.
6. Patterson, C., World Alzheimer Report 2018: the state of the art of dementia research: new frontiers. Alzheimer's Disease International (ADI): London, UK, 2018.
7. Ezzati, A., et al., Optimizing Machine Learning Methods to Improve Predictive Models of Alzheimer's Disease. Journal of Alzheimer's Disease, (Preprint): p. 1-10.
8. Jack Jr, C.R., et al., Age-specific population frequencies of cerebral β-amyloidosis and neurodegeneration among people with normal cognitive function aged 50–89 years: a cross-sectional study. The Lancet Neurology, 2014. 13(10): p. 997-1005.
9. Aguilar, C., et al., Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment. Psychiatry Research: Neuroimaging, 2013. 212(2): p. 89-98.
10. Bhagwat, N., et al., An artificial neural network model for clinical score prediction in Alzheimer disease using structural neuroimaging measures. Journal of psychiatry & neuroscience: JPN, 2019. 44(4): p. 246.
11. Mangialasche, F., et al., Alzheimer's disease: clinical trials and drug development. The Lancet Neurology, 2010. 9(7): p. 702-716.
12. Wee, C.Y., et al., Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. Human brain mapping, 2013. 34(12): p. 3411-3425.
13. Chincarini, A., et al., Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. Neuroimage, 2011. 58(2): p. 469-480.
14. Wolz, R., et al., Nonlinear dimensionality reduction combining MR imaging with non-imaging information. Medical image analysis, 2012. 16(4): p. 819-830.
15. Coupé, P., et al., Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease. NeuroImage: clinical, 2012. 1(1): p. 141-152.
16. Lee, G., et al., Predicting Alzheimer's disease progression using multi-modal deep learning approach. Scientific reports, 2019. 9(1): p. 1952.
17. Liu, C.-F., et al., Using Deep Siamese Neural Networks for Detection of Brain Asymmetries Associated with Alzheimer's Disease and Mild Cognitive Impairment. Magnetic resonance imaging, 2019.

13- Budget Details: (in Rials)

List of Expenses				Cost (amount in Rials)
Research Personnel and Partners' Compensation	**Name**	**Hours required**		
	1-			
	2-			
	3-			
	4-			
	5-			
	6-			
	7-			
	8-			
			Subtotal:	
Equipment and instruments (non expendable)	**Equipment and Model No.**	**Manufacturer**	**Quanti ty**	
	1-			
	2-			
	3-			
	4-			
	5-			
	6-			
			Subtotal:	
Equipment (expendable), lab animals and Materials	**Items**	**Manufacturer or Provider**	**Quanti ty**	
	1-			
	2-			
	3-			
	4-			
	5-			
	6-			
	7-			
	8-			
	9-			
	10-			
	11-			
	12-			
	13-			
	14-			
	15-			
Lab Tests and Services (specify)	**Laboratory Name**	**Number of Tests**		

Travel	Destinations	Transport Mode	Travelling	Number of trips	
	1-				
	2-				
	3-				
	4-				
Books, Copy and Print	Specify:				
Communication (phone, web etc)	Specify:				
Other Expenditures	1- 2- 3- 4- 5- 6- 7-				
				Grand Total:	

Important Note: For year 2014

- The maximum funding which may be assigned to **M.Sc.** theses is **25,000,000 Rials.**
- The maximum funding which may be assigned to **PhD** theses is **75,000,000 Rials.**
- The maximum funding which may be assigned to **Undergraduate** theses is **15,000,000 Rials.**

14- Are you going to receive any financial assistance/budget from other sources for this research? (if yes, specify from where and how much)

Project Manager(s) Affirmation:

Full Name:	Signature	Date
1-		
2-		
3-		